



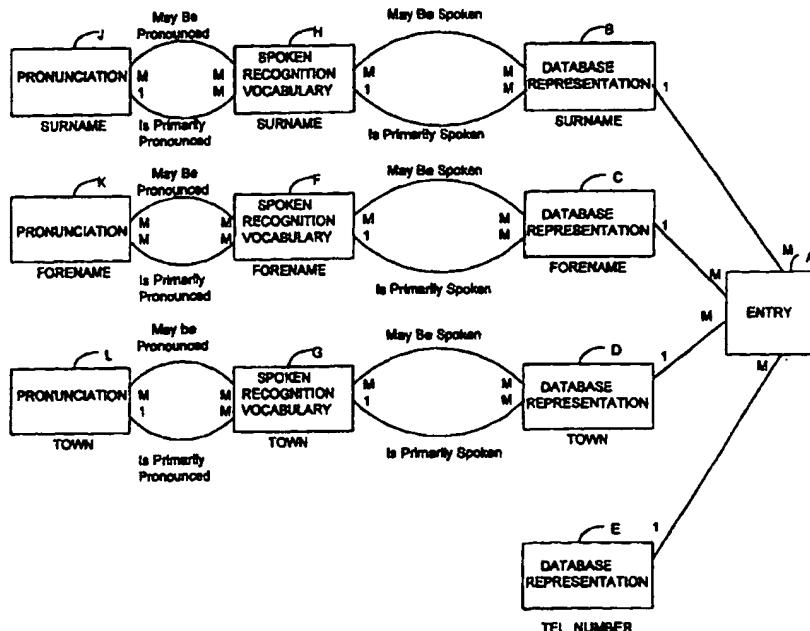
## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification <sup>6</sup> :	A1	(11) International Publication Number:	WO 97/28634
H04M 3/50, G06F 17/30, G10L 3/00		(43) International Publication Date:	7 August 1997 (07.08.97)
(21) International Application Number:	PCT/GB97/00233	(74) Agent:	HIBBERT, Juliet, Jane, Grace; BT Group Legal Services, Intellectual Property Dept., 8th floor, 120 Holborn, London EC1N 2TE (GB).
(22) International Filing Date:	27 January 1997 (27.01.97)	(81) Designated States:	AU, CA, CN, JP, KR, MX, NO, NZ, SG, US, European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).
(30) Priority Data:	9601925.2 31 January 1996 (31.01.96) GB	Published	<i>With international search report.</i>
(60) Parent Application or Grant			
(63) Related by Continuation			
US Filed on	659,526 (CIP) 5 June 1996 (05.06.96)		
(71) Applicant ( <i>for all designated States except US</i> ):	BRITISH TELECOMMUNICATIONS PUBLIC LIMITED COMPANY [GB/GB]; 81 Newgate Street, London EC1A 7AJ (GB).		
(72) Inventors; and			
(75) Inventors/Applicants ( <i>for US only</i> ):	ATTWATER, David, John [GB/GB]; 154 Cavendish Street, Ipswich IP3 8BG (GB). OLSEN, Paul, Andrew [GB/GB]; 65 Alexandra Road, Ipswich, Suffolk IP4 2RN (GB). BRIDGEMAN, Seamus, Aodhain [IE/IE]; Telecom Ireland Software, Merrion House, Merrion Road, Dublin 4 (IE). WHITTAKER, Steven, John [GB/GB]; 53 Bristol Road, Ipswich, Suffolk IP4 4LP (GB).		

## (54) Title: DATABASE ACCESS

## (57) Abstract

A method and apparatus for accessing a database system, said database system comprising a database containing entries each comprising a plurality of fields which contain machine representations of items of information pertaining to the entry, the said representations forming a first vocabulary; output means responsive to machine representations falling within a second vocabulary of such representations to generate signals representing the machine representations; and input means operable to receive signals and to produce machine representations falling within a third vocabulary of such representations. The method of accessing the database system comprises (i) generating, in accordance with a defined relationship between the first vocabulary and the third vocabulary, for each representation produced by the input means, one or more representations according to the first vocabulary; (ii) identifying database entries containing the generated representation; (iii) examining each representation or combination of representations which is contained in a selected field or combination of fields of the identified entries to identify distinguishable one(s) of those representations or combinations, a distinguishable representation or combination being one which, when translated in accordance with the defined relationship into representations of the second vocabulary, differs from every other such distinguishable representation or combination when similarly translated; and (iv) controlling the output means to generate an output including at least one word or combination of words which correspond(s) to one of the distinguishable representations or combinations.



The diagram illustrates the relationships between three main types of entities: PRONUNCIATION, SPOKEN RECOGNITION VOCABULARY, and DATABASE REPRESENTATION. These entities are interconnected through various fields such as SURNAME, FORENAME, and TOWN. The PRONUNCIATION entity is shown in three different forms (SURNAME, FORENAME, TOWN) and is linked to the SPOKEN RECOGNITION VOCABULARY entity via arrows labeled 'May Be Pronounced' and 'Is Primarily Pronounced'. The SPOKEN RECOGNITION VOCABULARY entity is also linked to the DATABASE REPRESENTATION entity via arrows labeled 'May Be Spoken'. The DATABASE REPRESENTATION entity is further connected to an ENTRY entity and another DATABASE REPRESENTATION entity (labeled E2). Finally, the E2 entity is connected to a TEL NUMBER entity.

***FOR THE PURPOSES OF INFORMATION ONLY***

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AM	Armenia	GB	United Kingdom	MW	Malawi
AT	Austria	GE	Georgia	MX	Mexico
AU	Australia	GN	Guinea	NE	Niger
BB	Barbados	GR	Greece	NL	Netherlands
BE	Belgium	HU	Hungary	NO	Norway
BF	Burkina Faso	IE	Ireland	NZ	New Zealand
BG	Bulgaria	IT	Italy	PL	Poland
BJ	Benin	JP	Japan	PT	Portugal
BR	Brazil	KE	Kenya	RO	Romania
BY	Belarus	KG	Kyrgyzstan	RU	Russian Federation
CA	Canada	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	KZ	Kazakhstan	SG	Singapore
CH	Switzerland	LJ	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovakia
CM	Cameroon	LR	Liberia	SN	Senegal
CN	China	LT	Lithuania	SZ	Swaziland
CS	Czechoslovakia	LU	Luxembourg	TD	Chad
CZ	Czech Republic	LV	Latvia	TG	Togo
DE	Germany	MC	Monaco	TJ	Tajikistan
DK	Denmark	MD	Republic of Moldova	TT	Trinidad and Tobago
EE	Estonia	MG	Madagascar	UA	Ukraine
ES	Spain	ML	Mali	UG	Uganda
FI	Finland	MN	Mongolia	US	United States of America
FR	France	MR	Mauritania	UZ	Uzbekistan
GA	Gabon			VN	Viet Nam

- 1 -

## DATABASE ACCESS

The present invention relates to a database access particularly, though not exclusively, employing speech recognition input and synthesised speech output.

5 International patent publication number WO94/14270 describes a mechanised directory enquiry system in which a caller is first prompted to speak the name of the city required. The word spoken is then recognised and the word with the highest confidence level is selected as being the word spoken by a user. The caller is then prompted to speak the name of the sought party. When a 10 satisfactory confidence level is obtained, a database is accessed and the number articulated to the caller. If the confidence level fails to meet a preferred confidence level, the caller is prompted to spell all or part of the location or the name. If more than one match between the spoken input and the database is found, the user is asked to confirm each match one by one until a confirmed match 15 is found. If no such match can be located, the automatic processing is terminated.

European patent application publication no. 433964 relates to a system which uses a text input. First of all an input word representing the surname is matched with the entries. If a match is found that is "comparable" but not exactly the same as the input, the initial characters of the input and the database entry are 20 compared. If these match, a record of the data entry is made. The system then compares the required titles and the personal names required. The most likely entry is provided to the user.

US patent no. 5204894 relates to a personal electronic directory in which the names of the entries are stored in the user's voice and associated numbers are 25 input using either a multitone (DTMF) telephone keypad or a spoken input. When a user needs to access the directory, the user speaks the required name and the directory system compares the first word of the input with the stored words and provides all possibilities in sequence to the user until the user confirms one.

In all of the prior art systems discussed above, the systems provide to the 30 user identified entries in a database in a sequential manner until a user confirms the data entry as being that required.

According to one aspect of the present invention there is provided a database access apparatus comprising:

- 2 -

- (a) a database containing entries each comprising a plurality of fields which contain machine representations of items of information pertaining to the entry, the said representations forming a first vocabulary;
- (b) announcement means responsive to machine representations falling 5 within a second vocabulary of such representations to generate audio signals representing spoken announcements;
- (c) input means operable to receive signals and to produce machine representations thereof falling within a third vocabulary of such representations;
- (d) translation means defining a relationship between the first vocabulary 10 and the second vocabulary and between the first vocabulary and the third vocabulary; and
- (e) control means operable
- 15 (i) to generate, in accordance with the defined relationship, for each representation produced by the input means, one or more representations according to the first vocabulary;
- (ii) to identify database entries containing the generated representations;
- (iii) to examine each representation or combination of representations which is contained in a selected field or combination of fields of the identified entries to identify distinguishable one(s) of those representations 20 or combinations, a distinguishable representation or combination being one which, when translated in accordance with the defined relationship into representations of the second vocabulary, differs from every other such distinguishable representation or combination when similarly translated; and
- 25 (iv) to control the announcement means to generate an announcement including at least one word or combination of words which correspond(s) to one of the distinguishable representations or combinations.

The included word output by the announcement means may be in any suitable form e.g. the included word may represent a whole word, a spelt word or 30 alphanumerics.

In another aspect the invention provides a method of speech recognition comprising

- (a) generating at least one announcement requiring a response;

- 3 -

- (b) recognising the response(s);
- (c) identifying database entries containing fields matching the recognised responses;
- (d) in the event that the number of such entries exceeds a predetermined limit, generating an announcement containing at least one word corresponding to a selected field of an identified entry for a positive or negative response;
- 5 (e) upon receipt of a positive response, identifying database entries which contain fields matching the recognised responses and whose selected fields match the said word; and
- 10 (f) repeating steps (d) and (e) at least once.

Some embodiments of the present invention will now be described, by way of example, with reference to the accompanying drawings in which:

Figure 1 is an Entity Relationship Diagram showing an example of  
15 translations between phonetic, spoken and database representations;

Figure 2 is a block diagram of apparatus according to the invention;

Figure 3 is a flow chart illustrating the operation of the apparatus of Fig. 1;

Figure 3a is a flow chart illustrating an alternative operation of the apparatus of Figure 1;

20 Figure 4 is a flow chart illustrating the process of identifying distinguishable tuples;

Figure 5 is an Entity Relationship Diagram showing an example of the translations between phonetic, spoken, spelt and database representations.

A voice interactive apparatus will be described, which generates questions  
25 to a user and recognises the user's responses in order to access the contents of the database. A database of names, addresses and telephone numbers, as might be used for an automated telephone directory enquiry system, will be used as an example. Firstly, however, some basic concepts will be discussed which will be of value in understanding the operation of the apparatus.

30 The database will be supposed to contain a number of entries, with each entry containing a number of fields each containing an item of information about the entry; for example the forename, surname, location and telephone number of the person to whom the entry refers. A set of fields from one entry is here referred to as a tuple, viz. a combination of N fields (when N = 1, 2 or 3 the terms single,

- 4 -

tuple and triple respectively are used). A complete entry is thus a tuple as also is a smaller set of fields extracted from one entry; thus a set of forename/surname pairs taken from the example database forms a set of extracted duples.

The items of information stored in the database fields may be in any 5 convenient representation; generally this description will assume the use of a text representation such as, for the surname Jonson, character codes corresponding to the letters of the name, but with a stylised representation for some fields; for example one might, for geographical locations, identify several distinct places having the same name with different representations - e.g. Southend1, Southend2 10 and Southend3 for three place in England called Southend.

The words used in the dialogue between the apparatus and the user to represent field contents are conceptually distinct from the database representations and represent for each field a **spoken vocabulary**. If the database representations are text then there will be some overlap between them and the 15 spoken vocabulary, but even then it may be desired to take account of the fact that the user might use, to describe an item of information, a different word from that actually contained in the database field; that is, some words may be regarded as **synonyms**.

Finally one needs also to note that more than one **pronunciation** may be 20 associated with a word (homonyms), and conversely more than one word may have the same pronunciation (homophones).

These concepts are illustrated in Figure 1 which is an "Entity Relationship Diagram", where we see a need for translation between representations as one moves from left to right or right to left. Box A represents a set of database entries. 25 Box B represents a set of unique surnames, which have a 1:many relationship with the entries -i.e. one surname may appear in many entries but one entry will contain only one surname. Boxes C, D and E correspond to sets of representations of forenames, towns and telephone numbers, where similar comments apply. Box F represents the spoken vocabulary corresponding to forenames i.e. the set of all 30 words that are permitted by the apparatus to be used to describe this field. This can differ from the database vocabulary (or, even if it is the same, may not have a 1:1 correspondence with it) to take account of aliases such as synonyms, for example an abbreviated form of a forename such as Andy or Jim may be

- 5 -

considered to have the same meaning as the full forms of Andrew and James. Two connecting paths are shown between boxes C and F, corresponding to a preferred form for the spoken vocabulary word and to alternative forms which "may possibly" be used.

- 5        Similarly, Box G represents the spoken vocabulary corresponding to town names. Here again the possibility of aliasing arises since often a large town may contain smaller places or districts within it. For example, Ipswich is a town in the county of Suffolk, England. Nearby is a small district called Kesgrave. A person living in Kesgrave might have his address recorded in the database either as  
10 Ipswich or as Kesgrave. Similarly an enquirer seeking the telephone number of such a person might give either name as the location. Thus Ipswich and Kesgrave may be regarded as synonymous for the purposes of database retrieval. Note however that this geographical aliasing is complex: Ipswich may be regarded as synonymous with another local village such as Foxhall, but Kesgrave and Foxhall  
15 are not synonymous because they are different places.

Box H represents, for completeness, a spoken vocabulary for surnames, though there is probably little scope for synonyms for this field.

Box J represents a pronunciation vocabulary for surnames, to take account of homophones and homonyms. For example the surname Smith is  
20 generally pronounced with a short "i" as in the English word "pith", whilst the name Smythe is pronounced with a long "i" as in "lithe". Smyth, on the other hand, may be pronounced either way. Other instances of varying pronunciation may arise, for example due to variations in regional accents. Again, "primary" and "may be" links are shown, for reasons to be explained later.

25       Boxes K and L represent pronunciation vocabularies for forenames and geographical names respectively.

Figure 2 is a block diagram of an apparatus for conducting a dialogue. An audio signal input 1 is connected to a speech recogniser 2, whilst an audio signal output 3 is connected to a speech synthesiser 4. A control unit in the form of a  
30 stored-program controlled processor 5 controls the operation of the recogniser and synthesiser and also has access to a program memory 6, a working memory (RAM) 7, a database 8, a spoken vocabulary translation table 9 and a pronunciation table 10. The audio inputs and outputs are connected for two-way communication - perhaps via a telephone line - with a user.

- 6 -

The database 8 is assumed to contain telephone directory entries, as discussed above, in text form. The spoken vocabulary translation table 9 is a store containing word pairs consisting of a directory representation and a spoken vocabulary representation, e.g., for the Ipswich example,

5	Database representation	Spoken representation
	IPSWICH	IPSWICH
	IPSWICH	KESGRAVE
	IPSWICH	FOXHALL
10	KESGRAVE	KESGRAVE
	KESGRAVE	IPSWICH
	FOXHALL	FOXHALL
	FOXHALL	IPSWICH

15 (If desired any word used as a database representation which has a 1:1 correspondence with, and is the same as, a spoken vocabulary word may be omitted from the table, since no translation is required). The translation table 9 has a separate area for each type of field and may be accessed by the processor 5 to determine the database representation(s) corresponding to a given vocabulary 20 word and *vice versa*. If desired (or if the database representations are not in text form) all items may be translated.

The pronunciation table 10 is a store containing a look-up table (and, if desired, a set of rules to reduce the number of entries in the look-up table) so that the processor 5 may access it (for synthesis purposes or for identifying 25 homophones) to obtain, for a given spoken vocabulary word, a phonetic representation of one or more ways of pronouncing it, and, conversely (for recognition purposes), to obtain, for a given phonetic representation, one or more spoken vocabulary words which correspond to that pronunciation. A separate area for each type of field may be desirable.

30 The operation of the apparatus is illustrated in the flow-chart of Figure 3 - which is implemented as a program stored in the memory 6. The first steps involve the generation, using the synthesiser, of questions to the user, and recognition of the user's responses. Thus in steps 100, 104, 108 the processor 5 sends to the synthesiser 4 commands instructing it to play announcements

- 7 -

requesting the user to speak, respectively the surname, forename and town of the person whose telephone number he seeks. In steps 102, 106 and 110 the processor sends to the recogniser 2 commands instructing it to recognise the user's responses by reference to phonetic vocabularies corresponding to those fields. The recogniser may access the translation table 9,10 to determine the vocabularies to be used for each recognition step, or may internally store or generate its own vocabularies; in the latter case the vocabularies used must correspond to those determined by the table 9,10 (and, if appropriate, the database) so that it can output only words included in the phonetic vocabulary.

10 The recogniser is arranged so that it will produce as output, for each recognition step, as many phonetic representations as meet a predetermined criterion of similarity to the word actually spoken by the user. (The recogniser could of course perform a translation to spoken vocabulary representations, and many recognisers are capable of doing so). It is possible that the recogniser may indicate that the

15 word actually spoken by the user is too dissimilar to any of the phonetic representations in Table 10 and indicate this to the processor 5. Preferably the recogniser also produces a "score" or confidence measure for each representation indicating the relative probability or likelihood of correspondence to the word actually spoken. The preliminary steps 100 - 110 will not be discussed further as

20 they are described elsewhere; for example reference may be made to our co-pending International patent application no. PCT/GB/02524.

In the following text, steps will be described which involve the matching of a number of scored tuples against database entries. From these matching entries a scored set of unique (or distinguishable) tuples are derived which

25 correspond to a different set of (possibly overlapping) fields to the tuples used for the match.

Following step 110 the processor 5 has available to it, for each of the three fields, one or more phonetic representations deemed to have been recognised. What is required now is a translation to spoken vocabulary representations - i.e. the translation illustrated to the left of Figure 1. Thus in step 30 112 the processor accesses the table 10 to determine, for each word, one or more corresponding spoken vocabulary representations, so that it now has three sets of spoken vocabulary representations, one for each field.

- 8 -

The score for each spoken vocabulary representation is the score for the phonetic representation from which it was translated. If two phonetic representations translate to the same vocabulary representation, the more confident of the two scores may be taken. This is a specific example of the 5 generalised matching process described above where the matching set of singles are pronunciations and the derived set of singles are spoken vocabulary items.

In step 114, the processor 5 now performs a translation to database representations - i.e. the translation illustrated in the centre of Figure 1 - using the table 9 to determine, for each spoken representation, one or more corresponding 10 database representations, so that it now has three sets of database representations. Scores may be propagated as for the earlier translation. The database representations represent a number of triples (the actual number being the product of the number of representations in each of the three sets). The score for a triple is typically the product of the scores of the individual representations of 15 which it is composed. At step 116, the processor generates a list of these triples and passes it to the database which returns a count K of the number of database entries corresponding to these triples. If (step 118) this number is zero, then the processor in step 120 sends a command to the synthesiser to play an announcement to the effect that no entry has been found, and terminates the 20 program (step 122). Alternatively other action may be taken such as transferring the user to a manual operator.

If there are entries, then in step 124 the full entry tuples which matched at step 116 are retrieved in turn from the database to determine whether there are three or fewer distinguishable entries. The tuples are retrieved in order of 25 likelihood, most likely first. It is possible that more than one tuple may share the same score. In this case an arbitrary ranking may be selected between them or a priori knowledge may be used to determine the ranking. As the tuples are retrieved, an assessment is made as to whether they represent three or fewer distinguishable entries. The meaning of "distinguishable" and the method of its 30 determination will be explained presently. Once a count of four is reached the test is terminated. If (step 126) the number of distinguishable entries is three or fewer, then in step 128 the processor retrieves these entries from the database and forwards them to the synthesiser 4 which reads them to the user in confidence

- 9 -

order, highest first, using the tables 9, 10 for translation from database representation to primary phonetic representation.

If there are more than three distinguishable entries then the process enters an iterative confirmation phase in which an attempt is made to identify lists of  
5 extracted tuples which contain three or fewer distinguishable tuples, and to offer the tuples in turn to the user for confirmation. In this example the tuples are the duple corresponding to the name (i.e. forename + surname), and the single corresponding to the town. Note that, although the case in this example, it is not in principle necessary that the constituent words of these tuples correspond to  
10 fields for which the user has already been asked.

In step 130 a check is made as to whether the name duples have already been offered for confirmation; on the first pass the answer will always be "no", and at step 132 a list of extracted name duples is prepared from the list of triples. The name duples from the list are examined in similar fashion to that of the triples  
15 in step 124 to determine whether there are three or fewer distinguishable duples. (If desired the number of non-identical database representation duples in the list may be counted, and if this exceeds a predetermined limit, e.g. 30 the detailed examination process may be skipped (to step 144)). If there are three or fewer distinguishable duples, (step 134) then each of the scored duples are translated  
20 into a single primary phonetic representation and fed to the synthesiser in confidence order in step 136 so that the synthesiser speaks the question (e.g.) "is the name John Smith? please answer yes or no" one at a time with the recogniser forwarding the reply to the processor (138) for testing for "yes" or "no". If the user replies "yes", then, in step 140:

- 25 (a) the surname and forename fields are marked "confirmed" so that further offering of them for confirmation is bypassed by the test at step 130;  
(b) all members of the list of triples, other than those which are related to the confirmed duple (see below), are deleted.

The process may then recommence from step 124.

- 30 If a user replies "no" then the corresponding members in the list of triples are deleted. Which ones are deleted depends upon the defined relationship between the phonetic representations and the database representations, as chosen by the system designer. For instance, if the user is asked "Is the name John Smith?" and the user replies "no", all members of the list of triples including John

- 10 -

and Smith/Smyth/Smythe may be deleted or only those members including John and Smith/Smyth may be deleted, it having been decided by the system designer that Smythe is always pronounced differently to Smith or Smyth.

Equally, if the user is asked "Is the name Dave Smith?" and the user 5 replies "no", the members including Dave Smith may be deleted and the user asked "Is the name David Smith?"

If (step 142) the user has answered no to all the offered tuples, this is considered a failure and the process it terminated via steps 120 and 122.

If in the test at step 134 the number of distinguishable name duples is too 10 large for confirmation, or at step 130 on a second or subsequent pass the name confirmation has already occurred, and assuming (step 144) the town name has not yet been offered for confirmation, then a town name confirmation process is commenced, comprising steps 146 to 154 which are in all respects analogous to the steps 132 to 142 already described.

15 If these processes fail to reduce the number of distinguishable entries at the test 126 then the process eventually terminates with an announcement 156 that too many entries have been found for a response to be given. Alternatively, a further procedure may follow in which one or more further questions are asked (as in step 100) to obtain information on further fields.

20 This process shown in Figure 3 from step 116 onwards has, for clarity, been described in terms of confirmation of a duple and a single. A more generalised algorithm might proceed as follows.

Start: If there are no database entries still active:

Give "none" message.

25 Finish algorithm.

Jump: If there are three or less distinguishable database entries:

Offer them.

Finish algorithm.

If there are more than three distinguishable database entries, then:

30 Do the following for successive prioritised fields or combinations of fields that have not already been confirmed until no such fields remain:

If for this there is a tuple list with 3 or less distinguishable tuples then:

- 11 -

Attempt to confirm this list.  
If positive confirmation, confirm it and go to JUMP.  
If negative confirmation give "wrong entry" message:  
go back to "do the following":

5

In a prioritised list, get the next vocabulary which may be asked.

If there is an un-asked and un-confirmed vocab remaining:

Ask for it.

Goto start of algorithm.

10 If not:

Give "too many" message.

Finish algorithm

An alternative process for the whole of the enquiry process is shown in  
15 Figure 3a. This process proceeds as follows:

Start

If there are no database entries still active (300):

give "none" message (301)

finish algorithm

20 If there are three or less distinguishable database entries (302):

offer them (303)

finish algorithm

If there are more than three distinguishable database entries (302),  
consider each of a prioritised list of fields or combinations of fields that  
25 have not yet been confirmed (304):

If there is a tuple list with three or less distinguishable tuples

Attempt to confirm this list (308)

If positive confirmation (309):

go to start

30 If negative:

give wrong entry message (310)

finish algorithm

If not:

- 12 -

Consider next field or combination of fields.

If no tuple lists with three or less distinguishable tuples:

In a prioritised list, if there is a remaining vocabulary which remains un-asked and un-confirmed (305):

- 5            ask for it (307)  
              go to "start"

if not:

give "too many" message (306)  
finish algorithm

10          In the above procedures, it is required to examine a list of tuples in database representation to determine how many distinguishable tuples there are. The tuple in question may be an entire database entry (as in step 124 above), it may be an extracted tuple containing representations from two (or more) fields (as in step 132) or it may be an extracted single (as in step 146).

15          Two representations are considered indistinguishable if:

- (a) they are identical; or
- (b) they translate to identical spoken vocabulary words (e.g. they are synonyms or are geographically confused); or
- (c) they translate to spoken vocabulary words which are homophones (i.e. those words translate to identical phonetic representations).

Two tuples are considered indistinguishable if every field of one tuple is indistinguishable (as defined above) from the corresponding field of the other tuple.

Equally two representations are considered distinguishable if:

- (a) they are not identical; and
- (b) they do not translate to identical spoken vocabulary words (e.g. they are not synonyms or geographically confused); and
- (c) they do not translate to spoken vocabulary words which are homophones (i.e. the words do not translate to identical phonetic representations).

Suppose that we have a list of tuples in database representation where the first tuple in the list is D(1) and the tuple currently occupying the n'th position in the list is D(n) where n=1,...,N, there being N tuples in the list. Each tuple consists of M fields, designated d, so that the m'th field of tuple D(n) is d(n,m) - i.e. D(n) = {d(n,m)}, m = 1,...,M. Preferably the list is ordered by score; i.e. the tuple having the highest confidence is D(1), the next D(2) and so on.

- 13 -

The process to be described is illustrated in the flowchart of Figure 4 and involves taking the first tuple from the list, and comparing it with the tuple below it in the list to ascertain whether the two are distinguishable. If they are not, the tuple occupying the lower position is deleted from the list. This is repeated until all  
5 tuples have been examined. The same steps are then performed for the tuple now occupying the second position in the list, and so on; eventually every tuple remaining in the list is distinguishable from every other. If desired, the process may be terminated as soon as it is certain that the number of distinguishable tuples exceed that which can be handled by subsequent steps (i.e., in this example, 3).

10 In Figure 4, i points to a tuple in the list and j points to a tuple lower down the list. I is the number of tuples in the list. In step 200, i is initialised to 1, and I is set to N, and in step 202 D(i) is read from the database. Step 204 sets j to point to the following tuple and in step 206 D(j) is read. A field pointer m is then initialised to 1 in step 208 and this is followed by a loop in which each field of the  
15 two tuples is taken in turn. Field m of tuple D(i) is (step 210) translated, with the aid of the table 9, into one or more spoken vocabulary words s1(a) where a = 1,...A and A is, effectively, the number of synonyms found. The spoken vocabulary word(s) a1(a) is/are then translated (212) with the aid of the table 10 into a total of B phonetic representations p1(b) (b = 1,...B). B is the number of  
20 such representations, i.e. A multiplied by the number of homophones. Analogous steps 214, 216 perform a two-stage translation of the corresponding field of D(j) to produce one or more phonetic representations p2(d) (d = 1,...D).

In step 218, each of the phonetic representations p1(b) is compared with each of the representations p2(d) (i.e. BD comparisons in total). If equality is not  
25 found in any of these comparisons, then the two tuples are considered distinguishable. If (step 226) j has not reached the last tuple in the list, it is incremented (228) prior to reading a further tuple in a repeat of step 206; otherwise the tuple pointer i is tested at step 230 as to whether it has reached the penultimate member of the list and either (if it has not) is incremented (232) prior  
30 to a return to step 202, or (if it has) the process ends. At this point, the list now contains only mutually distinguishable tuples - I in number - and thus the result k is set to I in step 233 prior to exit from this part of the process at step 234.

If on the other hand the comparison at 218 indicates identity between one of the phonetic representations generated for one field of one tuple and one of the

- 14 -

phonetic representations generated for the same field of the other tuple then it is necessary to increment m (step 236) and repeat steps 210 to 218 for a further field. If all fields of the two tuples have been compared and all are indistinguishable then this is recognised at step 238 and the tuples are deemed to 5 be indistinguishable. In this case, the lower tuple D(j) is removed from the list and I is decremented so that it continues to represent the number of tuples remaining in the list (steps 240, 242). j is then tested at step 244 to determine whether it points beyond the end of the (now shortened) list and if not a further tuple is examined, continuing from step 206. Otherwise the process proceeds to step 230, 10 already described.

Each time step 232 increments i to point to a tuple, it is known that there are at least i tuples which will not be removed from the list by step 240. Thus at this point i can be tested (step 246) to see if it has reached 3, and if so the process may be interrupted, k set to 4, and thence to the exit 234.

15 In order to clarify the relationship between the algorithm of Figure 4 and the steps of Figure 3 or 3a, it should be mentioned that:

- (a) the algorithm represents the execution of step 124, with the list at the conclusion of Figure 4 being used to access (from the database) the entries to be offered in step 128;
- 20 (b) the algorithm represents the execution of step 132, with the list at the conclusion of Figure 4 representing the list of name duples (in database representation) to be offered to the user in step 136;
- (c) the algorithm represents the execution of step 146, with list at the conclusion of Figure 4 representing the list of towns to be offered to the user in 25 step 150;
- (d) the algorithm represents the execution of step 302.

It remains to explain the removal which occurs in steps 140 and 154 in Figure 3. Taking step 140 as an example, the principle followed is that: where the user has confirmed a tuple (in this case a duple) which is one of a pair (or group) of 30 tuples deemed indistinguishable, then this is considered to constitute confirmation also of the other tuple(s) of the pair or group. For example, if the list of name duples contains:

Dave Smith

- 15 -

**David Smyth**

and these are considered by step 132 to be indistinguishable, only one entry (for example the first "Dave Smith") is offered to the user for confirmation in step 136. Which tuple is presented is determined according to a choice made by the system  
 5 designer. However, if the user says "yes", then in step 140, all tuples containing "Dave Smith" and all tuples containing "David Smyth" are retained.

Whilst this could be done using the results of the translations performed in step 132, we prefer to proceed as follows. Each field p of the confirmed duple in phonetic representation (i.e. the one generated in step 136) is translated using the  
 10 tables 9, 10 into one or more database representation. All duples represented by combinations of these representations are to be confirmed - i.e. any of the list of triples which contains one of these duples is retained, and the other triples are deleted.

It is perhaps worth clarifying the relationship between the Entity  
 15 Relationship Diagram of Figure 1 and the processes set out in Figures 3 and 4. In these processes, translations occur from database representation to spoken vocabulary representation to phonetic representation (i.e. right to left in Figure 1) and in the opposite direction, viz. from phonetic representation to spoken vocabulary representation to database representation (i.e. left to right in Figure 1).  
 20 The existence of alternative paths in the diagram (e.g. may be spoken/is primarily spoken) implies a choice of translation routes. For synthesis, the "primarily spoken" routes would normally be used; for other purposes, variations are possible according to one's desire to include or exclude synonyms or homophones in the translation. Different routes are tabulated below, with an example set of routes  
 25 for forenames. Other mappings may be used.

Direction	Used in steps	Description	Typical Route for forenames
input to database	112, 114	Mappings used to convert a recognition result into all possible database representations	MayBePronounced..../ MayBeSpoken..... (i.e. include synonyms and homophones)

- 16 -

database to output	124, 132, 146 302, 304	mappings used to decide on distinguishable database representations	IsPrimarilySpoken..../ MayBePronounced.... (i.e. results in synonyms but not homophones being included in the resulting list of distinguishable tuples)
database to output	124, 132, 146 302, 304	mappings used to decide on distinguishable database representations	IsPrimarilySpoken/ IsPrimarilyPronounced (i.e. excludes homophones from output but includes synonyms and homonyms)
database to output	136, 150 128, 303, 308	mappings used for output of result e.g. synthesis	IsPrimarilySpoken/ IsPrimarilyPronounced (i.e.provides a primary output form for each database representation)
output to database	140, 154 308	mappings used to confirm an output pronunciation back into database representation	MayBePronounced/ MayBeSpoken (i.e. include both synonyms and homophones) /
output to database	140, 154, 308	mappings used to exclude items rejected by user	Maybe Pronounced/ Is Primarily Spoken (i.e.excludes homophones but includes synonyms for subsequent database searches)

Thus when an input is received and machine representations representing the input signal are generated, the machine representations for the single input are converted to all possible data representations. To achieve this, the input is mapped to the spoken recognition vocabulary by the "May be pronounced" route. Thus all possible spoken representations of the input are identified. These spoken

- 17 -

vocabulary representations are then mapped onto all possible database representations which the spoken representations may represent (e.g. "May be spoken"). For example, say the forename "Dave" is input and the phonetic representations D AI V and M AI V are generated by the speech recogniser 2.

5 These phonetic representations are then converted to spoken vocabulary representations, for instance, "Dave" and "Mave". Each of these spoken vocabulary representations is then converted, by means of the store 9, into all possible database representations ("May be spoken") e.g. Dave, David, Mave, Mavis. The processor 5 then searches the database 8 for all entries including any  
10 of these entries in their forename field.

For all entries found the processor 5 then examines each representation contained in the selected field to identify distinguishable ones of those combinations. The distinguishable entries may then be presented to the user for confirmation. To achieve this, the processor 5, with reference to the spoken  
15 vocabulary store 10 and according to the defined relationships, translates the identified database representations into spoken vocabulary via the "may be spoken" route. Thus all entries of Dave are translated to "Dave" and "David", all entries for David are translated to "Dave" or "David", all entries for Mave are translated to "Mave" or "Mavis" and all entries of Mavis are translated to "Mave"  
20 or "Mavis". The processor 5 then translates the spoken vocabulary representations into phonetic representations ("may be pronounced") with reference to store 10. Thus the phonetic representations which represent how "Dave", "David", "Mave" and "Mavis" are pronounced are determined as D AI V, D AI V I D, D AA V I D, M AI V and M AI VI S. These phonetic representations are  
25 then examined to identify distinguishable ones. For example, Dave and David are indistinguishable because they share at least one common pronunciation. However, Mave and Dave are distinguishable because they do not share any common phonetic representation. If two database representations are found to be indistinguishable, one of the representations is maintained and the other is  
30 discarded e.g. David may be selected over Dave and Mavis over Mave. This choice is determined by the system designer and stored in memory 6. The phonetic representation of the most probable of "David" and "Mavis" is presented by the processor 5 to the synthesiser 4 to a user using the "Is primarily spoken"/"is primarily pronounced" relationship.

- 18 -

Note that in practice the stores 9, 10 may contain separate "tables" for each mapping.

Figure 1 shows a situation in which the vocabulary of the announcement means e.g. the speech synthesiser is the same as the vocabulary of the input means e.g. the speech recogniser. However this is not always so. For instance, it should be noted that spellings may also be used as an alternative input and/or confirmation medium to spoken forms. The techniques required for spelling are directly analogous to spoken forms. Figure 5 corresponds to Figure 1 with the inclusion of spelling (illustrated by box M) for town names (spelling may also be provided for surnames and forenames although, for simplicity, these mappings have not been shown in Figure 5). Translations of "may be spelt" and "is primarily spelt" must be provided in addition to the spoken recognition.

If spellings are to be used during recognition and/or confirmation then for all the routes mentioned above with reference to Figure 1, "Spelt" is substituted for 'Pronounced' and the algorithms all still apply.

It should also be mentioned that spoken or spelt input or output is not essential - since the considerations concerning the offering and confirming still arise. For example, keypad input could be used, which has ambiguity problems owing to the allocation of more than one letter to each button of a telephone keypad. In this case a further vocabulary - of keypad input codes - is required, with "May be keyed ...." translations analogous to the pronunciation and spelling translations described above.

The machine representations of the input vocabulary and the database may be generated according to the same technique, for instance the database entries may be stored in text form and the input also be in text form, with the machine representations of the database and the output being generated according to a different technique e.g. a spoken output.

Confusion may arise if a user is presented with an announcement which includes a synonym of the actual word said by the user. For instance, say a user asks for "Dave Smith" and the system generates an output as follows: "Did you say David Smith?" In order to avoid this confusion, a check may be carried out to ensure that the word corresponding to the identified distinguishable database entry corresponds also to the word recognised by the input means.

- 19 -

CLAIMS

1. A database access apparatus comprising:
  - (a) a database containing entries each comprising a plurality of fields which contain machine representations of items of information pertaining to the entry, the said representations forming a first vocabulary;
  - (b) announcement means responsive to machine representations falling within a second vocabulary of such representations to generate audio signals representing spoken announcements;
  - 10 (c) input means operable to receive signals and to produce machine representations thereof falling within a third vocabulary of such representations;
  - (d) translation means defining a relationship between the first vocabulary and the second vocabulary and between the first vocabulary and the third vocabulary; and
- 15 (e) control means operable
  - (i) to generate, in accordance with the defined relationship, for each representation produced by the input means, one or more representations according to the first vocabulary;
  - (ii) to identify database entries containing the generated representations;
  - 20 (iii) to examine each representation or combination of representations which is contained in a selected field or combination of fields of the identified entries to identify distinguishable one(s) of those representations or combinations, a distinguishable representation or combination being one which, when translated in accordance with the defined relationship into representations of the second vocabulary, differs from every other such distinguishable representation or combination when similarly translated; and
  - (iv) to control the announcement means to generate an announcement including at least one word or combination of words which correspond(s) to one of the distinguishable representations or combinations.
- 30 2. Apparatus according to claim 1 wherein the control means is operable to control the announcement means to generate successive announcements, each of which includes at least one word or combination of words which correspond(s) to

- 20 -

one of the distinguishable representations or combinations, the control means being operable to control the announcement means to output the announcements in sequential confidence order, the first announcement including at least one word or combination or words which correspond(s) to the most likely distinguishable  
5 representation or combinations.

3. An apparatus according to claim 1 or 2 in which the control means is operable, in step (iv), for the or each distinguishable representation or combination, to generate, using the translation means, from the distinguishable representation or  
10 combination, one representation or combination in the second vocabulary and to transmit this to the announcement means.
4. An apparatus according to claim 1 or 2 in which the control means is operable, in step (iv), for the or each distinguishable representation or combination,  
15 to transmit to the announcement means one representation or combination in the second vocabulary which corresponds, in accordance with a relationship defined by the translation means, to the distinguishable representation or combination and which has already been generated in step (iii).
- 20 5. An apparatus according to any of claims 1 to 4 in which the control means is operable in step (iv) to generate an announcement requesting confirmation of the included word or combination and is further arranged, in operation:
  - (v) upon receipt of a confirmatory response, to generate from a representation or combination in the second vocabulary, which corresponds  
25 to the included word(s), one or more representations or combinations according to the first vocabulary and to identify the database entries or entry which contains such a representation or combination in the selected field(s).
- 30 6. An apparatus according to any one of the preceding claims in which the input means is a speech recogniser operable to receive audio signals.
7. An apparatus according to claim 6 in which the second and third vocabularies are identical.

- 21 -

8. An apparatus according to any preceding claims wherein the first and third vocabularies are identical.
- 5 9. An apparatus according to any preceding claim in which at least one of the selected field(s) is a field in which, in step (ii), a generated representation was found, and in which a word included at step (iv) is a word which corresponds to a representation generated by the input means.
- 10 10. Apparatus according to any preceding claims further comprising an intermediate vocabulary and the translation means defines the relationships between the first and the intermediate vocabulary; the second and the intermediate vocabulary; and the third and the intermediate vocabulary.
- 15 11. A method of accessing a database system, said database system comprising a database containing entries each comprising a plurality of fields which contain machine representations of items of information pertaining to the entry, the said representations forming a first vocabulary; announcement means responsive to machine representations falling within a second vocabulary of such 20 representations to generate audio signals representing spoken announcements; and input means operable to receive signals and to produce machine representations falling within a third vocabulary of such representations;
- the method of accessing the database system comprising:
- (i) generating, in accordance with a defined relationship between the first 25 vocabulary and the third vocabulary, for each representation produced by the input means, one or more representations according to the first vocabulary;
- (ii) identifying database entries containing the generated representations;
- (iii) examining each representation or combination of representations which is contained in a selected field or combination of fields of the identified entries to 30 identify distinguishable one(s) of those representations or combinations, a distinguishable representation or combination being one which, when translated in accordance with the defined relationship into representations of the second vocabulary, differs from every other such distinguishable representation or combination when similarly translated; and

- 22 -

(iv) controlling the announcement means to generate an announcement including at least one word or combination of words which correspond(s) to one of the distinguishable representations or combinations.

5 12. A method according to claim 11 further comprising, in step (iv), controlling the output means to output one or more announcement(s) in sequential confidence order, the first announcement including at least one word or combination of words which corresponds to the most likely distinguishable representation.

10 13. A method according to Claim 11 or 12 further comprising:  
(a) generating at least one announcement requiring a response;  
(b) recognising the response(s);  
(c) identifying database entries containing fields matching the recognised responses;

15 (d) in the event that the number of such entries exceeds a predetermined limit, generating an output containing at least one word corresponding to a selected field of an identified entry for a positive or negative response;  
(e) upon receipt of a positive response, identifying database entries which contain fields matching the recognised responses and whose selected fields match the said  
20 word; and

(f) repeating steps (d) and (e) at least once.

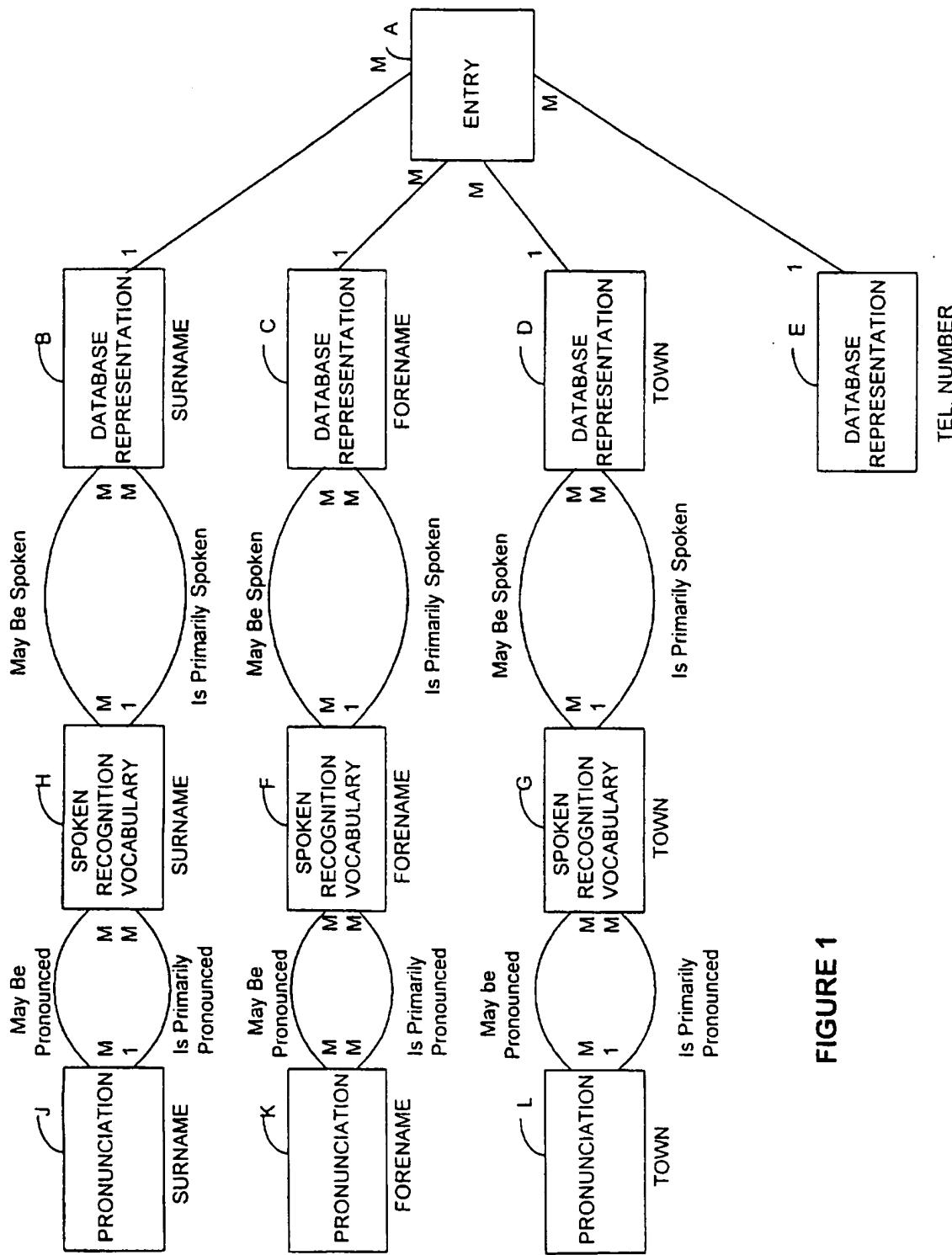
14. A method according to claim 11 or 12 comprising:

(a) generating at least one announcement requiring a response;  
25 (b) recognising the response(s);  
(c) identifying database entries containing fields matching the recognised response(s);  
(d) in the event that the number of such entries is below or equal to a predetermined limit, generating an output presenting one or more of the entries  
30 each containing one or more of the field(s) of the matching entries, and exit;  
(e) in the event that the number of such entries exceeds a predetermined limit, for a particular field or selection of fields, examining a distinguishable representation or selection of representations, which is contained in the particular field or combination of fields of the identical entries;

- 23 -

- (f) in the event that the number of such distinguishable representations or combinations of representations is above a predetermined limit, repeat step (e), selecting another field or selection of fields, according to a pre-determined order, that has not already been considered until no such fields remain to be considered;
- 5 and
- (g) generating at least one output requiring a response that has not already been requested or confirmed;
- (h) recognising the response(s);
- (i) repeat step (c) at least once;
- 10 (j) in the event that the number of such distinguishable representations or combinations of representations is below or equal to a predetermined limit, generating an announcement containing at least one word corresponding to the selected field(s) of an identified entry for a positive or negative response;
- (k) upon receipt of a positive response, identifying database entries
- 15 which contain fields matching the recognised responses and whose selected fields match the said words or combination of words; and
- (l) repeating steps (d) and (e) at least once;
- (m) upon receipt of a negative response for all such word or words, exiting the database accessing method.

17



## FIGURE 1

2/7

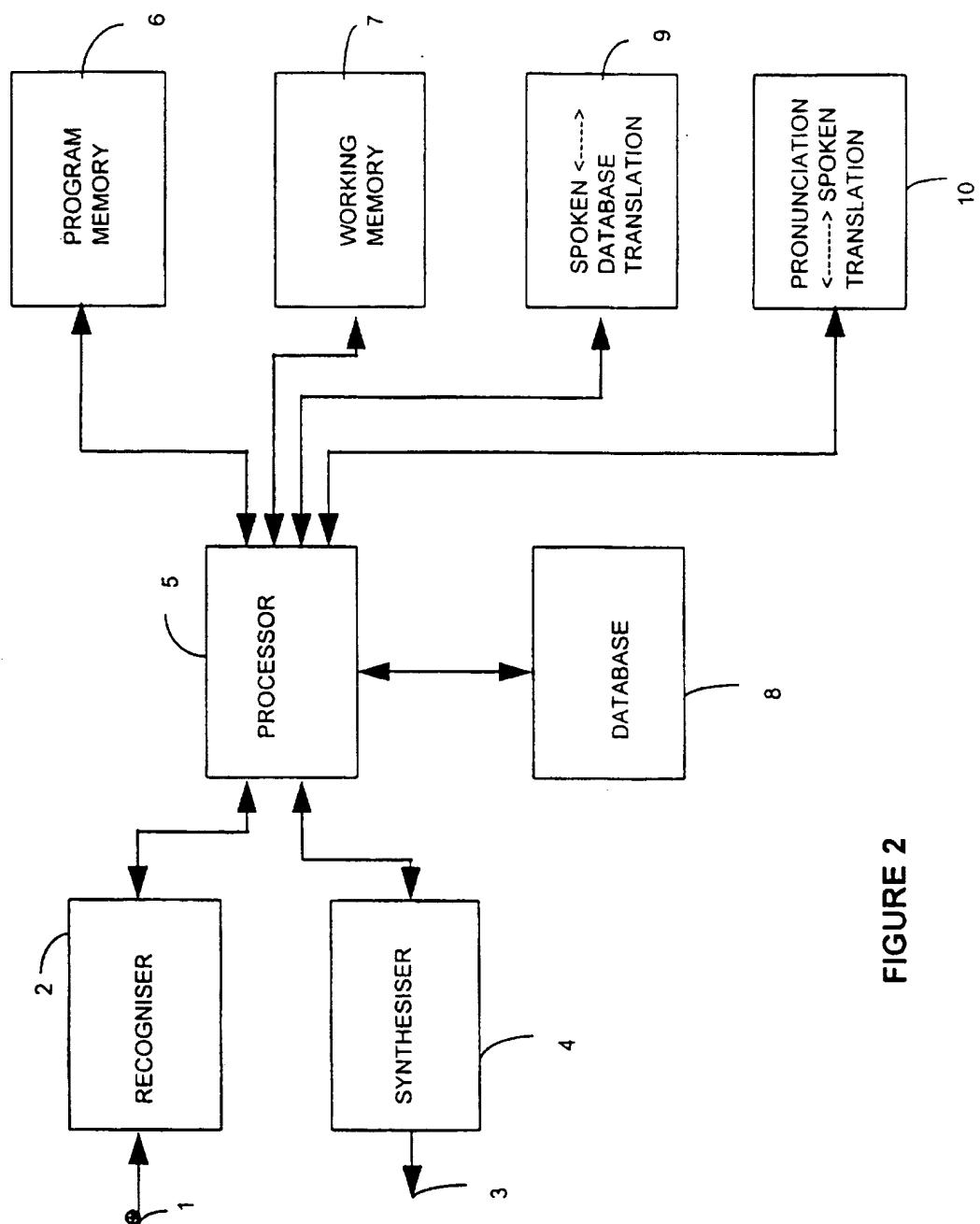
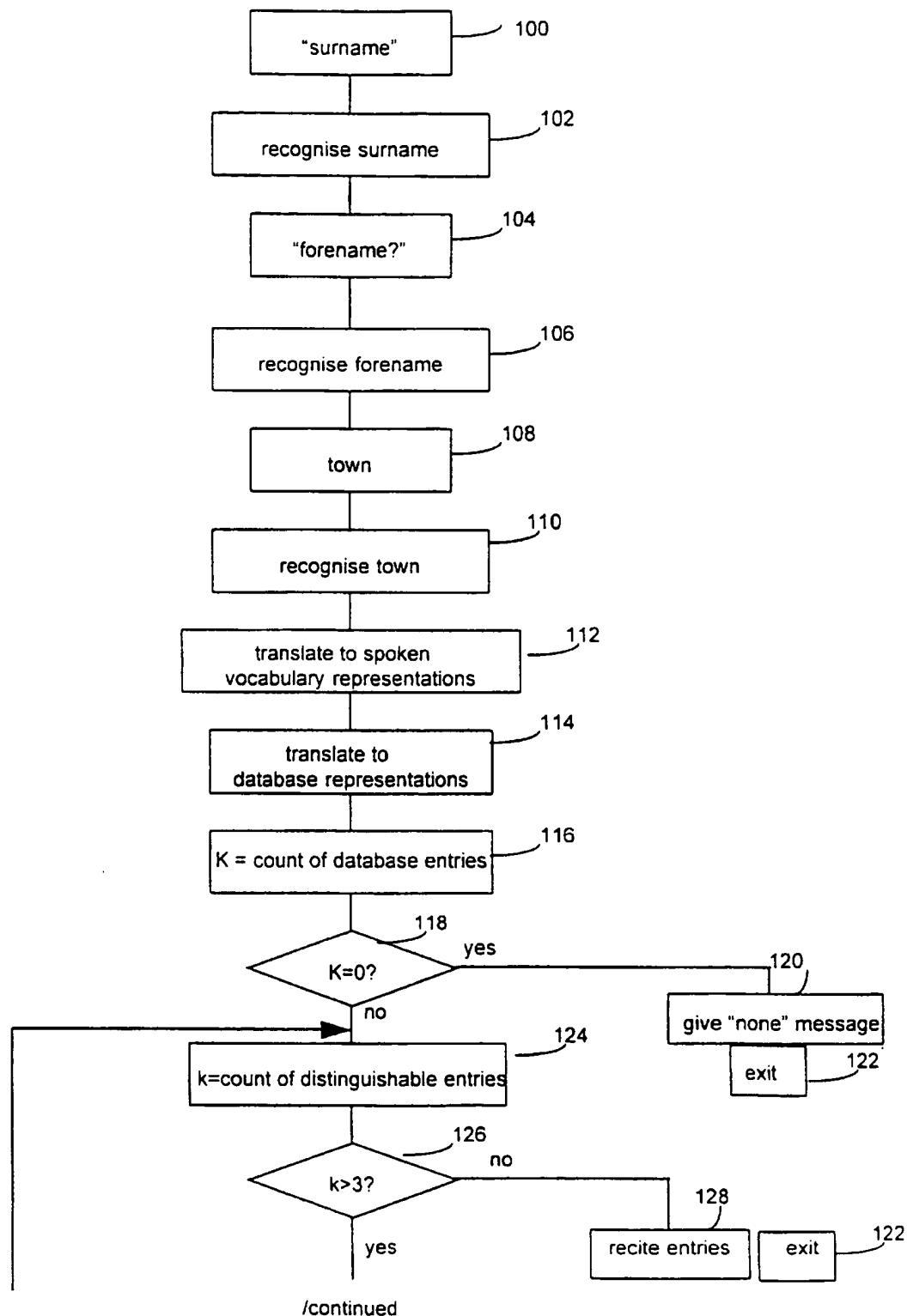


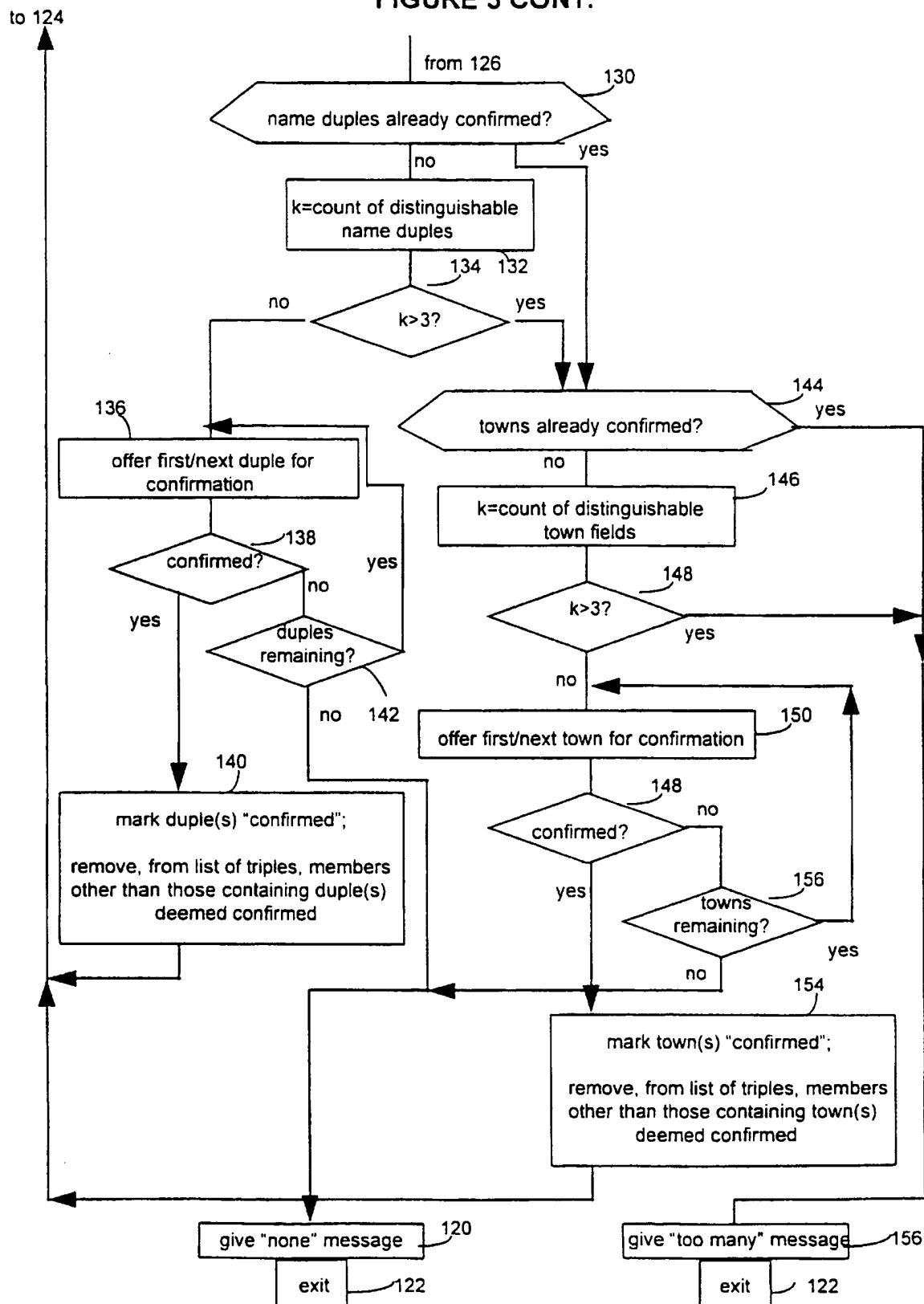
FIGURE 2

3/7  
**FIGURE 3**

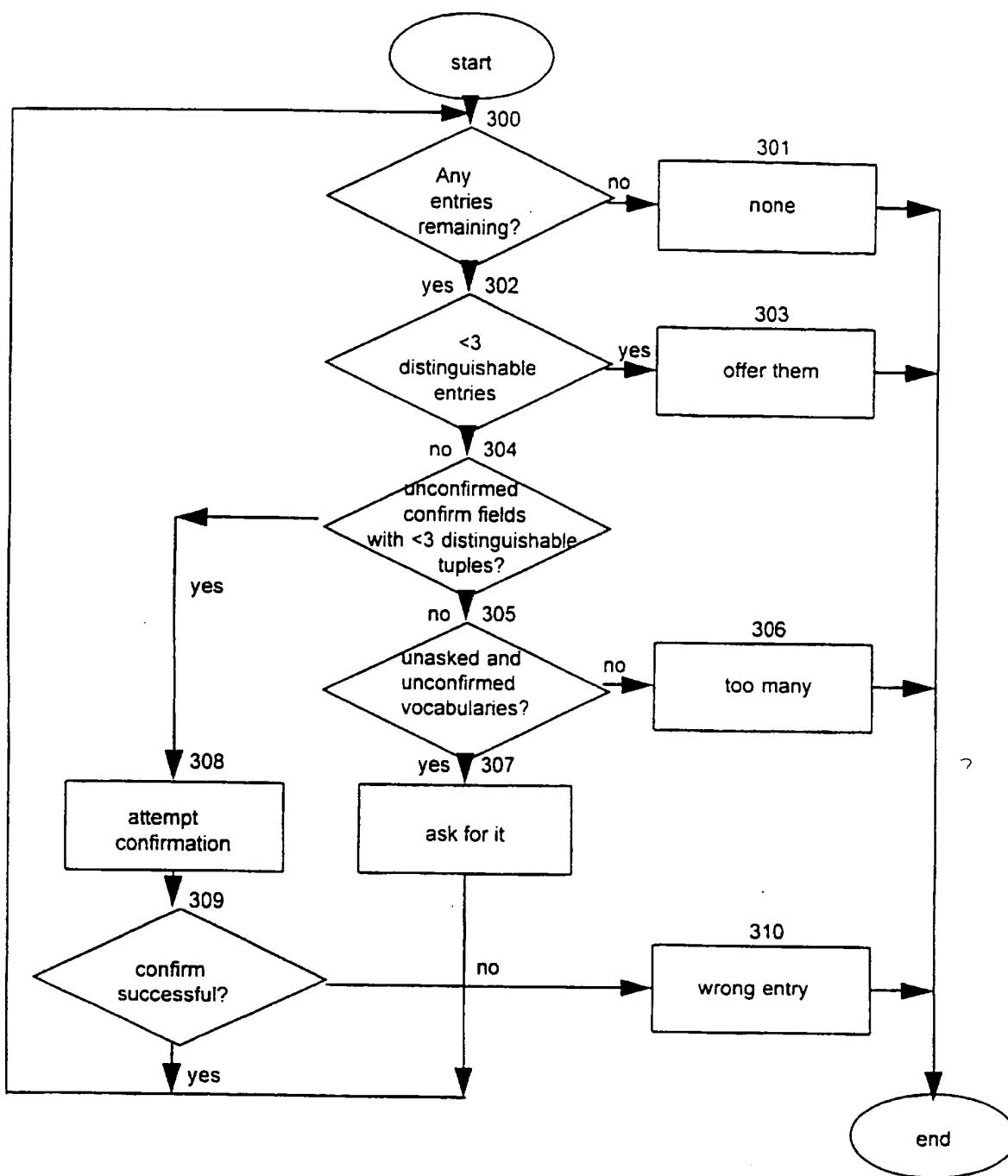


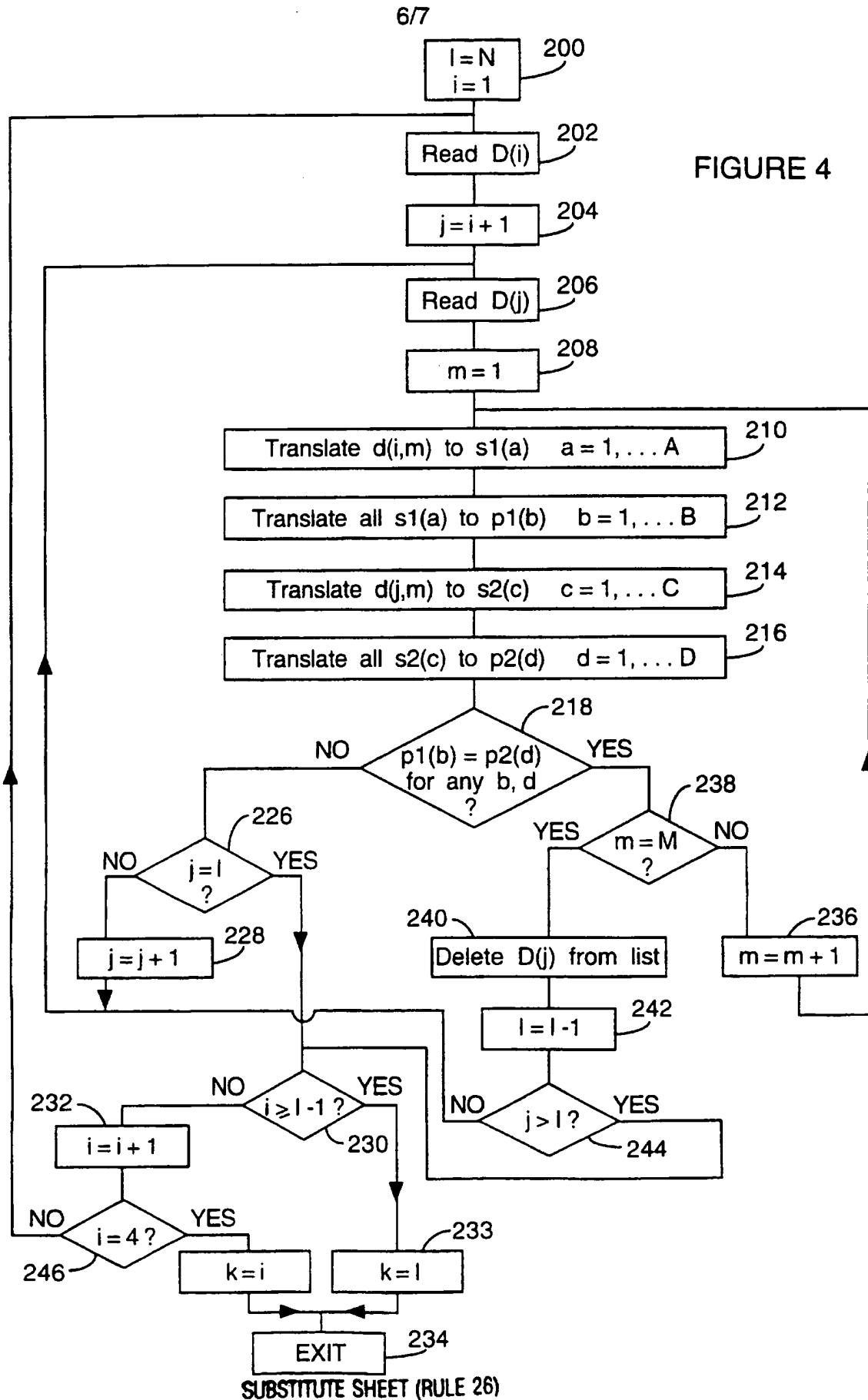
/continued

4/7



5/7  
**FIGURE 3A**





7/7

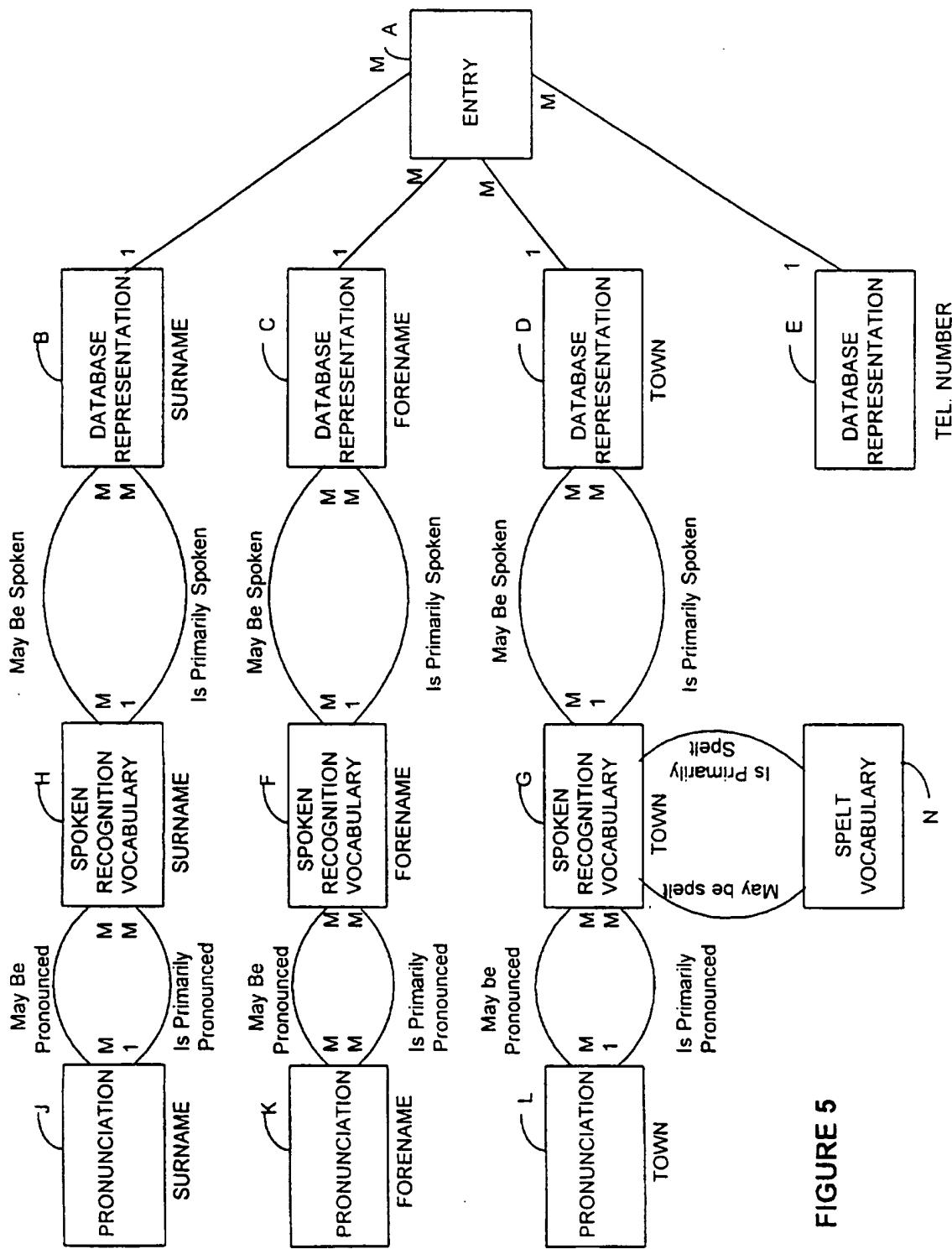


FIGURE 5

# INTERNATIONAL SEARCH REPORT

International Application No  
PCT/GB 97/00233

**A. CLASSIFICATION OF SUBJECT MATTER**  
**IPC 6 H04M3/50 G06F17/30 G10L3/00**

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
**IPC 6 H04M G06F G10L**

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	BT TECHNOLOGY JOURNAL, vol. 14, no. 1, January 1996, IPSWICH SUFFOLK (GB), pages 177-186, XP000579339 D.J. ATTWATER ET AL: "ISSUES IN LARGE-VOCABULARY INTERACTIVE SPEECH SYSTEMS" see paragraph 4.5 see paragraph 5.1 see paragraph 7.1 --- WO 94 14270 A (BELL ATLANTIC) 23 June 1994 cited in the application see abstract -----	1,11
A		1,11



Further documents are listed in the continuation of box C.



Patent family members are listed in annex.

\* Special categories of cited documents :

- 'A' document defining the general state of the art which is not considered to be of particular relevance
- 'E' earlier document but published on or after the international filing date
- 'L' document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- 'O' document referring to an oral disclosure, use, exhibition or other means
- 'P' document published prior to the international filing date but later than the priority date claimed

- 'T' later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- 'X' document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- 'Y' document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
- '&' document member of the same patent family

1

Date of the actual completion of the international search	Date of mailing of the international search report
28 April 1997	13.05.97
Name and mailing address of the ISA  European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl. Fax (+31-70) 340-3016	Authorized officer  Vandevenne, M

**INTERNATIONAL SEARCH REPORT**

Information on patent family members

International Application No

PCT/GB 97/00233

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9414270 A	23-06-94	AU 5803394 A	04-07-94